DOCUMENT RESUME

ED 196 953 TM 810 104

AUTHOR Fleishman, Allen I.
TITLE Investigation and Comparison of Alternative Weighting Systems.
PUB DATE [78]
NOTE 51p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Error of Measurement; Factor Analysis; Least Squares Statistics; Mathematical Models; *Multiple Regression Analysis; *Predictive Measurement
IDENTIFIERS *Monte Carlo Studies; *Weighted Grading Systems

ABSTRACT

A Monte Carlo study was performed to test under what conditions and to what degree various alternatives would be superior to multiple linear regression (MLR). The criteria used was the mean squared error of prediction of individual scores and the mean squared error of estimation of the regression weights. Populations were simulated containing either a two or four factor latent factor structure and high, medium, low, or wide range of communalities. Sampling was stimulated with 16, 64 or 256 'subjects'. Forty types of weights were examined. In all conditions an alternative weighting method was superior to MLR. The structural methods, principal components or factor analytic regression, were generally the best for most conditions. Validity weights were generally superior to both MLR and unit weights. The shrunken weights consistently dominated MLR. As much as a 64- fold reduction in errors was found. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED196953

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Investigation and Comparison of Alternative Weighting Systems

by

Allen I. Fleishman

Lederle Laboratories

Statistical Design and Analysis

Pearl River, New York 10965

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. Fleishman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Alternative Weighting Systems

TM 810104

Abstract

A Monte Carlo study was performed to test under what conditions and to what degree various alternatives would be superior to MLR. The criteria used was the mean squared errors of prediction of individual scores and the mean squared error of estimation of the regression weights. Populations were simulated containing either a two or four factor latent factor structure and high, medium, low, or wide range of communalities. Sampling was simulated with 16, 64 or 256 'subjects'. Forty types of weights were examined. In all conditions an alternative weighting method was superior to MLR. The structural methods, principal components or factor analytic regression, were generally the best for most conditions. Validity weights were generally superior to both MLR and unit weights. The shrunken weights consistently dominated MLR. As much as a sixty four fold reduction in errors was found.

Keywords: Biased Estimators, Factor Analytic Regression, Factor Analytic Weights, Mean Square Error, MLR, Monte Carlo Study, OLS, Principal Component Regression, Shrunken Weights, Unit Weights, and Validity Weights.

Introduction

Much research has been done to predict a criterion from many predictors. The criterion is usually predicted by a linear combination of the predictors. In prediction one is primarily interested in the accuracy of the prediction of individual scores.

Composites are also used for theory building as well. In theory building, one is primarily interested in the accuracy of the weights and not necessarily in the accuracy of the prediction of individual subjects.

The most frequent method used to predict scores and parameters is Multiple Linear Regression (MLR). By this technique a vector of regression weights is generated.

By assuming the "classical linear regression model" (Kmenta, 1971, 348), the estimated MLR regression weight vector will have the following properties

1. minimize errors in the sample,
2. be unbiased,
3. have the smallest Mean Square Error (MSE) of prediction across all other unbiased estimators,
4. be asymptotically unbiased, and
5. be a consistent estimator (Kmenta, 1971, 216).

If the assumption of the fixed X is relaxed to allow the predictors to form a random vector (e.g., X being vectors of predictor scores from randomly selected S's), then it can be shown that B will still be a maximum likelihood (Sampson, 1974) and least squares estimator (Kmenta, 1971, 299).

It is useful to consider the relative utility of each of these properties of MLR in more detail.

1. Minimize errors in the sample - A property of least square error estimators. This property implies that no other linear weight vector can predict Y as well as Y' = XB with $\sum(Y - Y')^2$ as the criterion.

In as much as this is true only in the sample and not necessarily true in the population, this property could be considered trivial. That is, in a sample one knows both the predictors as well as the criterion. Therefore in the sample, given both predictor and criterion one can predict the criterion perfectly. The ability to predict the criterion in a second sample (or the entire population) is not addressed by this property.

2. Unbiasness - A property of the best linear unbiased estimator. Given an infinite number of samples the mean of all these samples will be the vector of population weights. The estimated vector from any single sample will neither over nor under estimate the population vector.

An unbiased mean does not imply that any sample vector of weights will be close to the population parameters. It is possible that the weights are distant from the true parameters, but are randomly distant and overpredict as often as underpredict (i.e., unbiased). There may exist biased weights which on average may be slightly biased estimates of the population parameters, but nevertheless be more accurate, and have smaller squared error when predicting B_{pop} (Winer, 1978).

3. Smallest MSE of all unbiased estimators - This is a property of the best linear unbiased estimators. This property implies that in the population, errors of prediction will be the smallest of all other unbiased estimators.

This property is perhaps the most powerful for the applied data analyst. However it is only true across the class of unbiased estimators. As implied above it may be possible to find a smaller MSE estimator in the class of biased

estimators.

The smallest MSE across all other unbiased estimators property will not apply when the predictors are random variables and not fixed.

4. & 5. Asymptotically unbiased and consistant estimators - This is a property of maximum likelihood estimators. This property implies that as the sample size (N) becomes infinite, the estimate converges, unbiased, on the true population parameters, B_{pop}.

These properties neither informs us about how well the estimate performs in small samples, nor informs us what constitutes a large sample. It only applies in the case of an infinite sample size.

In sum, the MLR estimate is the best unbiased estimator across the domain of all other unbiased estimators. When N is very large it will converge on the true population parameters, B_{pop}.

In a classic paper by Stein (1960), it was demonstrated that there were estimators which uniformly dominated over the MLR estimator, i.e., expected to have equal or smaller MSE for all situations. Unfortunately there are many possible such estimators.

The purpose of this investigation was to determine which of the biased estimators were superior to MLR, as well as, what conditions are necessary for MLR to be superior to other methods. A number of alternatives to MLR have been suggested in the literature. These include unit, validity, principal components, factor analytic, and shrunken weights. These five alternative methods were examined.

Unit Weights

In this method the z scores are summed to yield a single predictor. Empirically it has been found that the unit weighting system has equal or superior predictive ability as compared to the MLR weights under many conditions (Claudy,

1972; Dawes & Corrigan, 1974; Dorans & Drasgow, 1978; Einhorn & Hogarth, 1975; Lawshe & Schmucher, 1959; Schmidt, 1971, 1972; Wesman & Bennett, 1959).

The general conclusion from Monte Carlo studies has been that when N is small one should use unit weights (Dorans & Drasgow, 1978; Schmidt, 1971, 1972). On the other hand it has been observed that if validities have wide variation (Claudy, 1972), or suppressors exist (Schmidt, 1971, 1972), or if the multiple correlation is very high (Eishorn & Hogarth, 1975), then unit weights may be superior only for extremely small N.

Wainer (1976) attempted to prove a theorem which stated that equal weights would have almost no loss in accuracy of prediction. However his theorem was shown to be less general than previously claimed (Laughlin, 1978; Pruzek & Frederick, 1978). Even so, the use of equal weights, which is quite insensitive to sampling fluctuations, may be very effective in many types of situations.

In sum, equal weights may be superior to MLR when 1) there is limited or poor information (e.g., small N or small R^2) or 2) the first centroid accounts for most of the intercorrelations among the predictors (Pruzak & Frederick, 1978). As suggested above, when N is large it is expected that MLR will prove superior with many types of data sets.

When a single battery is given to differentially predict many criteria, unit weights would obviously be inappropriate. That is, if unit weights were applied to the subscales of an intelligence test to predict job aptitudes, the predicted scores would be the same for each of the aptitudes considered. No differential prediction would be possible.

Validity Weights

Rather than weighting each score equally, this alternative weights each z score by its correlation with the criterion (i.e., its validity coefficient).

Examination of Claudy's work (1972) revealed that the validity weighting method (which he refered to as criterion correlation weights) was superior to unit weights when N was large. In addition, Claudy's tables revealed that the validity weights were superior in more populations than MLR when N was 160 (which was the largest N Claudy utilized).

The validity weights might be expected to converge on the true population weights, B_{pop}, under a number of conditions. One of the clearest conditions would be the case where the predictors are independent of one another. Another way of describing this case, would be where zero factors is an adequate model to describe the interrelationships among the predictors. It might be theorized that the validity weights would be superior to unit weights in a number of circumstances. In the case where the predictors were not of equal importance, and their importances were proportional to their correlations with the criteria, the validity weights would be superior to unit weights if the other factors were constant. While the unit weights rely only on the direction of each variable (either stated a priori or by the sign of the validity coefficient), the validity weights rely not only on the direction but also its magitude. As the magnitude is more sensitive to sample fluctuations than direction (especially if the direction is known a priori), the unit weights would be less sample dependent and would be superior in smaller samples. This result has been observed by Claudy (1972) and Dorans & Drasgow (1978).

In both unit and validity weights, the only parameters of interest are related to the validity coefficients. The interrelations among the predictor variables, standard deviations and means are ignored.

Principal Component and Factor Analytic Weights

The following two sets of methods, the principal component weights (Horst, 1941; Winer, 1978) and factor analytic weights (Horst, 1941; Lawley & Maxwell, 1973; Scott, 1966, 1969) assume that a smaller number of causal variables exist underlying the interrelationships of the variables. In both the principal component and factor analytic approaches, the regression weights can be obtained by the traditional MLR technique, where the covariances among the predictors (perhaps including the criterion) are approximated by a lower rank decomposition. Collectively the principal component and factor analytic techniques shall be referred to as the 'structural' methods, as both techniques approximate the structure of the interrelations of the variables.

It has been suggested that the factor analytic weights may be used to great benefit in significance testing (Scott, 1966). This suggestion has been empirically supported as it has been shown that the confidence interval will be much shorter (hence will yield a more powerful test) than the traditional interval. Using the random predictor model it was shown that the regression weights are asymptotically unbiased and normally distributed (Scott & Fleishman, 1976, 1978).

In a number of simulations, the principal component and factor analytic methods have consistently proven themselves superior to MLR. Leiman (1951) demonstrated that factor analytic weights, based on decomposing both predictors and criterion by the centroid method was superior to MLR in cross validation with real data in all but the largest sample (N = 200). Burket (1964) concluded that principal component weights, decomposing only the predictors, was superior to MLR in Mean Squared Error of the the estimate of the true regression weights, MSE(B). Hertzberg (1969) found that principal component on the predictors was maximally superior to MLR in cross validation when N was small. Pruzek, Sherry & Huba

(1975) used both a principal component and a factor analytic type weighting method, based on image factor analysis (Joreskog, 1962). They found, when factoring both the predictors and criterion, that both methods were generally quite superior to MLR. Principal components were found superior to factor analytic weights for high variable correlations; factor analytic weights were superior to principal component weights for low variable correlations; and principal component were superior to factor analytic weights when doublet factors exist with the criterion. In a study utilizing a large number of methods, Dempster, Schatzoff & Wermuth (1977) found the principal component method, on the predictors only, to be superior to MLR in both MSE(B) and Mean Squared Error of prediction, MSE(Y), when one extracts as many components as the foward selection F ratio > 1.0. Gunst & Mason (1977) also used a large number of alternative weighting methods. In their study principal component factoring the predictors and a technique similar to principal component, decomposing both predictors and criterion, (which they refer to as the latent root technique) were found best for multicollinear data. Dorans and Drasgow (1978) used both factor analytic and principal component weights decomposing only the predictor covariance matrix. They found the factor analytic weights only slightly superior to MLR. Principal components (extracting the correct number of factors) were superior to MLR across all levels of N used (up to N = 120) in predicting the regression weights with a congruence coefficient; and only slightly poorer than unit weights in cross validation to the population and superior to MLR except when N was large (N = 120). In an unpublished study this author foung that factor analytic weights, based on the image factor analysis (Joreskog, 1962), were superior to MLR in both MSE(B) and MSE(Y), but inferior to other factor analytic methods (i.e., using highest r as communality estimate).

Shrunken Weights

It is well known that MLR tends to overfit the data, hence overestimates the regression weights. Stein (1960) demonstrated that when there were more than two parameters to estimate, shrinking the MLR weights toward zero dominated MLR (i.e., MLR was an inadmissable statistic). Sclove (1968) had extended this work and developed a general algorithm which synthesized the shrunken estimates and preliminary significance testing. One of Sclove's weighting methods (Sclove, 1968, Eqn. 6), makes the shrinking constant monotonically proportional to the predictability of MLR [i.e., multiplying the shrinking constant by $(1 - R^2)/R^2$], and proportional to the number of subjects [i.e., shrinking proportional to $(n-2)/(N-n+2)$], where N was the number of subjects and n the number of parameters to be estimated. Therefore, greater shrinkage occurred when 1) R^2 was small, 2) N was small, and 3) the number of parameters to be estimated, n, was large.

In the study by Gunst & Mason (1977) the shrunken weights were always superior to MLR, except when the multiple correlation was very high, $R^2 = 0.9999$. For orthogonal data, the shrunken weight was superior to all methods, especially MLR, except for a ridge regression method. When a variable existed with considerable overlap with other variables, the shrunken weights were only slightly superior to MLR and poorer than the principal component and ridge regression methods. Dempster, Schatzoff, & Wermuth (1977), concluded "the relatively conservative STEINM [a shrunken estimator - AF] procedures has one advantage, in that it less ofter does worse than" MLR.

It is unfortunate that the results of many previous studies may be of limited generalizability. A number of studies were undertaken with natural data (Leiman, 1951; Lawshe & Schucker, 1959; Wesman & Bennett, 1959; Burket, 1964) or simulated sampling from natural correlation matrices (Schmidt, 1971, 1972). These two paradigms seldom allow researchers to generalize their findings to other data

sets. That is, most of these studies dealt with prediction of college grades (Leiman, 1951; Lawshe & Schucker, 1959; Wesman & Bennett, 1959; and Burket, 1964). Given that the same criteria and similar predictors were used in each, generalization to other type data sets would be difficult. In addition to this problem, by using an obtained data set, it is uncertain what parameters of the data are responsible for the results. That is, if two methods give different results, only a posteriori hypotheses may explain why (e.g., multicollinearity, outliers, number of factors). Therefore it is necessary to 1) investigate the effects of the alternative weighting procedures in more than one class of data, 2) vary the population parameters to be appropriate for a large class of representative populations, and 3) know a priori how the populations differ (i.e., to control the relevant parameters experimentally).

A second class of Monte Carlo studies utilized simulated populations (Herzberg, 1969; Claudy, 1972; Pruzak, Sherry, & Huba, 1975; Gunst & Mason, 1977; Demster, Schatzoff & Wermuth, 1977; and Dorans & Drasgow; 1978). One problem with simulated data is that it must allow parameters to vary in the range seen in applied literature. One study which illustrates the problems with the two types of Monte Carlo studies is that of Claudy (1972). He created eighteen populations, but was only able to explain differential conclusions a posteriori. Therefore it is unknown whether the populations he created simulate naturally occuring covariance matrices, or whether Claudy had experimental control of the population parameters.

Unfortunately the most comprehensive investigations to date generated data from some quite atypical populations (Gunst & Mason, 1971; Demster, Schatzoff & Wermuth, 1977). For example, the population multiple correlations in the Gunst & Mason study were 0.038, 0.500, 0.990, and 0.9999. Furthermore, in that study the predictors were either all orthogonal to one another, or were all orthogonal with

the exception of one predictor, which was a linear combination of three other predictors (their multicollinear case). Finally the fixed predictor - regression model - was utilized. The Demster et. al. study (1977) had very similar difficulties. Their population R^2 varied from 0.907 to 0.997, very unrepresentative of most social sciences multiple correlations. Multicollinearity was again induced by allowing two (of the six) variables to correlate 0.99. In addition the regression model was used. Both of these studies might apply to the ANOVA - fixed orthogonal predictor use of the general linear model, but would be inappropriate for general prediction.

Methodology

The present study attempted to determine under what conditions alternative weighting systems predict better than MLR. Since many of the relevant sampling distributions are currently unknown, Monte Carlo methodology was used to empirically simulate these distributions.

Forty weighting systems were used. They were: MLR, unit weights, validity weights, shrunken weights, principal component weights based on the predictors (Pri-X), principal component weights based on both the predictors and criterion (Pri-X:Y), classical factor analytic regression (CFAR) based on the predictors (F.A.-X), and CFAR based on both predictors and criterion (F.A.-X:Y). The latter four methods extracted from one to six factors.

The populations had one criterion and twelve predictors and varied on the number of underlying factors (2 or 4), communality level (high, medium, low, or wide) for each variable, and had a single replication. There were 100 replications from each population sampling either 16, 64 or 256 'subjects'.

The methodology of Tucker, Koopman, and Linn (1969) was selected to generate the population covariance matrices. Their procedure was based on the factor analytic model. For each set of population matrices (there was one replication), it was hypothesized that an investigator might attempt to select variables to measure (in varing degrees) the underlying variables.

It was desirable to vary the number of underlying latent variables within each population. Two and four variables were selected for this study. In the two variable case, a linear sum across the first two and last two columns of the factor loading matrix was performed. This procedure insured that the two and four factor cases would have parallel effects for the two and four factor populations.

In order to simulate the varying reliability of each variable, the degree of communality (or conversely, the degree of uniqueness) was varied. The degree of communality was set at four levels - high communality, $H^2 \sim U(.70 - 0.90)$; medium communality, $H^2 \sim U(.40 - 0.60)$; low communality, $H^2 \sim U(.10 - .30)$; and wide communality, $H^2 \sim U(.10 - 0.90)$. In order to insure comparability across communality levels, the four levels of communality were selected to be linear combinations of one another:

1) $h^2_{\text{medium, j}} = h^2_{\text{high, j}} - 0.3$ observed attribute j = 1,...,13

2) $h^2_{\text{low, j}} = h^2_{\text{medium, j}} - 0.3$ observed attribute j = 1,...,13

3) $h^2_{\text{wide, j}} = 4(h^2_{\text{low, j}}) - 0.3$ observed attribute j = 1,...,13

Therefore, for each population, there were two levels of factors and four levels of communalities. Thus there were eight population matrices per population replication. The means and variances of each variable was set at 0.0 and 1.0, respectively, within each of the populations.

The population regression weights and multiple correlation for all population 1 matrices are presented in Table 1.

*** Insert Table 1 about here ***

In the Tucker, Koopman & Linn method (1969), random noise was used to simulate varying levels of communality and the experimenter's ignorance of the variables. In order to determine the importance of such random perturbation on the model (i.e., to simulate the effect of different experimenter's choice of variables), the above procedure was replicated.

Most standard errors (e.g., $s_{\bar{x}}$, s_r, s_b) are functions of the number of observations (N). N was set at 16, 64, and 256 observations. The lower level was selected because it yielded very few degrees of freedom while allowing the matrix to be non-singular. The upper limit was selected since pilot work suggested a lack of discrimination among techniques in terms of MSE, if N were raised. Within each sample size and for each population, one hundred sample replications were generated. A sample replication (given the underlying population) was constructed by the methodology of Wijsman (1959) using Wishart (1928) matrices. Each sample replication generated a variance-covariance matrix and a vector of means.

Per each level of N there were 100 sample replications. Therefore, there were 4800 mean and covariance matrices of 12 predictors and criterion (two population replications X two levels of number of factors X four levels of communality X three levels of number of observations (N) X one hundred sample replications). A more detailed description of this methodology can be found in Fleishman (1978).

It was assumed that the investigator had no a priori knowledge of the mean or covariance structure (other than the assumption of the positive manifold). All weighting methods will be based on sample statistics. The following are the

alternative weighting methods used in this investigation.

Multiple Regression Weights:

Based on the normal equations, the vector of the regression weights is defined to be:

4) $b_{MLR} = C_{X,X}^{-1} C_{X,Y}$

where $C_{X,X}$ is the sample variance covariance matrix of the predictors, and

$C_{X,Y}$ is the vector of sample covariances of the predictors and the criterion.

Shrunken Weights:

The shrunken weights are

5) $b_{Shrunken} = sb_{MLR}$

where $b_{Shrunken}$ is the vector of Shrunken weights,

$s = [1 - c(1 - R^2)/(R^2)]$,

R^2 is the sample multiple correlation based on the sample MLR weights,

c is in the interval $0 < c < 2(n-2)/(N-n+2)$,

N is the number of Ss, and

n is the number of predictors (=12).

The constant c was selected such that it was in the middle of the optimal range.

6) $c = (n-2)/(N-n+2)$

Unit Weights:

Under the above assumptions (as an investigator would not typically be aware of the differences or lack of differences among means or variances) it was decided to utilize an equal weighting system which will apply equal weights to standardized scores to generate a composite. This composite was generated by:

7) $Z = s_Y X_1/s_1 + s_Y X_2/s_2 + \ldots + s_Y X_{12}/s_{12} + c$

where Z is the composite score,

s_Y is the sample standard deviction of the criterion Y,

X_i is the score for the predictor i,

s_i is the sample standard deviation of X_i, and

c is an additive contant.

However, as the composite may only be marginally related to Y, it was desirable to apply simple regression to this composite, Z. The weights used were

8) $b_{unit,i} = s_Y c_{ZY}/(s_i c_{ZZ})$

where $b_{unit,i}$ are the unit weights for predictor i,

c_{ZY} is the sample covariance of the composite Z, and the criterion Y, and

c_{ZZ} is the sample variance of the composite Z.

Validity Weights:

As with the unit weights, a composite T, was generated. Instead of being equally weighted, a predictor was weighted by its validity coefficient.

9) $T = c_{YX(i)}/s^2_i$

where $c_{YX(i)}$ is the sample covariance of the criterion, Y, and the i-th predictor, x_i.

As with the unit weights simple regression was used to predict the criterion.

10) $b_{Validity,i} = c_{YX(i)} c_{TY}/(s^2_Y c_{TT})$

where $b_{Validity,i}$ is the Validity weight for predictor i,

c_{TY} is the covariance of the composite T and the criterion, and

c_{TT} is the variance of composite T.

Factor Analytic and Principal Component Methods

Pilot work suggested that Image Factor Analysis or Factor Analysis of Rescaled Covariances (Joreskog, 1962) may not perform as well as other factor analytic methodologies. More traditional methods were selected. Two popular communality estimates were used: the multiple correlations and the highest r. The author also considered whether it might be more predictive to factor the entire matrix (predictors and criterion) or just factor the predictors. In addition it is not known how many factors to extract, in applied settings. Since there were only 12 predictors, an upper limit of six factors was used. Therefore, there were 2 communality estimating procedures x 2 levels of number of variables to factor x six levels of factors extracted.

The lower rank approximation to the predictor covariance matrix, $C^{-1}_{L;X,X'}$ was inserted into the traditional MLR formulea, Equation 4, for the two structural methods. The resultant vector was then restandardized as by Equation 8 or 10. The lower rank approximation was created by $AA' + U^2$ or AA', for the factor analytic or the principal component technique, respectively. While the factor analytic approximation to the covariance matrix is non-singular, the principal component approximation was singular. A generalized inverse was used for principal components.

11) $C^{-1}_{L;X,X} = V\Lambda^{-1}V'$

where V is the first m eigenvectors of the decomposition of $C_{X,X'}$ and

Λ is a diagonal matrix containing the first m eigenroots of the decomposition.

When decomposing both the predictors and the criterion, the lower rank approximation to the predictor covariance matrix and the lower rank approximation to the validity vector could have been inserted into the traditional MLR formulea, Equation 4, as above. However it was easier to approximate the full predictor-criterion covariance matrix, invert this covariance matrix (or form a generalized inverse, as by Equation 11 above), then generate the unstandardized weight matrix by the MLR formulea for the full matrix inverse:

12) $$w_{X:Y,X:Y} = (C^{Y,Y}_{L;X:Y,X:Y})^{-1} C^{Y,X}_{L;X:Y,X:Y}$$

where $w_{X:Y,X:Y}$ is a column vector of unstandardized weights based on the lower rank covariance matrix, $C_{L;X:Y,X:Y}$,

$C^{Y,Y}_{L;X:Y,X:Y}$ is a scalar in the row and column corresponding to the criterion in the inverse of the lower rank covariance matrix, and

$C^{Y,X}_{L;X:Y,X:Y}$ is the column vector containing the column corresponding to the criterion and the rows corresponding to the predictors of the inverse (or generalized inverse) of the lower rank matrix.

This weight vector would then be standardized as above.

In sum there were 40 weighting methods suggested, they were:

1 = MLR,

2 = Shrunken weights,

3 = Unit weights,

4 = Validity weights,

5 - 10 = High r factoring predictors only (F.A.-X),

11 - 16 = High r factoring predictors and criterion (F.A.-X:Y),

17 - 22 = Multiple correlation factoring predictors only (F.A.-X),

23 - 28 = Multiple correlation factoring predictors and criterion (F.A.-X:Y),

29 - 34 = Principal components factoring predictors only (Pri-X), and

35 - 40 = Principal components factoring predictors and criterion (Pri-X:Y).

The primary criterion to evaluate the efficacy of any weighting system is the Mean Squared Error (MSE) of prediction of individual scores [MSE(Y)] and the Mean Squared Error of estimate of the true regression weights [MSE(B)]. It was expected that logarithmic transformations of MSE would be more regular than the variable itself.

Results

In this experiment there were 192,000 observations for each dependent variable. Rather than address all the statistically significant effects, only those effects which had a partial omega square (Fleishman, in press) greater than 0.001 will be considered below. This cutoff point would correspond to an alpha level of approximately 1.0×10^{-40}. The formula for the partial omega square for any effect, w^2 is given by

13) $$w^2_{effect} = (SS_{effect} - df_{effect} MS_{error}) / ((SS_{effect} + SS_{error}) + MS_{error})$$

MSE(Y) - Errors of Prediction

A question addressed in this study was how often would an alternative weighting method be superior to (i.e., predict individual scores with less errors than) MLR. One alternative method utilized was principal components decomposing four factors of the predictor covariance matrix, Pri-X-4. These results are illustrated in Table 2.

If the percentage were 100%, the principal components weighting method would always have been superior to MLR. If the percentage were 50%, the principal components method would be found better than MLR as often as MLR would be superior to the principal component technique. The results from each cell were based on two hundred replications. The standard error of any one of the percentages would be largest when the true proportion was 50%. Therefore, the standard error of any cell proportion must be equal to or less than 3.5%.

*** Insert Table 2 about here ***

As can be seen in Table 2, the principal component method was always superior to MLR when N was small (N = 16). In every one of 1600 samples observed, the principal component method predicted with less error. As N increased, MLR converged to the true population weights. Although MLR did improve, the principal component technique was clearly superior. When N was moderate (N = 64), the principal component technique was superior 97.5% of the time. When N was large (N=256), the principal component technique was superior to MLR 86.69% of the time, across all levels of communality and factors. With respect to communality, when the variables had either medium or low communality levels the principal component technique was superior to MLR more than 98.5% of the time for each level of N. When N was large and the communality level was either high or wide, MLR was relatively competitive. Only in the wide communality, four factor population, with N = 256, was MLR found to be as good as the principal component technique.

In sum, when N was small or when the variables had medium or low correlations (communality levels from 0.10 to 0.50), the principal component technique virtually always had less errors in prediction than the multiple regression technique.

The above has addressed the percentage of times an improvement of any size could be made in prediction. The amount of improvement in prediction was another important consideration. If one examines the ratio of MSE(Y) for MLR to MSE(Y) for the principal component technique, (since both can be considered to be measured on a ratio level scale), a ratio can be formed. It is the author's contention that this ratio would indicate the proportion of errors in the MLR technique to the errors in the principal component technique (i.e., proportion of relative errors).

If this ratio were 1.00 the MSE(Y) of MLR would be equal to the MSE(Y) for principal components (i.e., the amount of squared error of the methods would be equal). If the ratio were 1.10, the MLR would have 10% more squared errors than the alternative (i.e., the principal component technique reduced the magnitude of squared errors by 10%). It should be noted that this ratio is independent of the original unit of measurement. The proportions of relative errors are presented in Table 3.

*** Insert Table 3 about here ***

As can be seen in Table 3, there apparently was a monotonically decreasing relationship between sample size and the relative superiority of Pri-X-4. When N was small (N = 16), the MLR technique had, on the average, more than five times more error than the principal component technique. When N was moderate (N = 64), improvements of 16% were found. When N was large (N = 256), slight improvements, 2%, were found by use of the principal component technique. When N was large, the communality or number of factors in the population apparently had little effect on the superiority of the principal component method, in which case the MLR and the principal component methods were approximately equal. When N was small, communalities had a larger effect. The two factor population showed greater improvement, for each level of N, than the more complex four factor population.

Following Dempster, Schatzoff, & Wermuth (1977), it was felt that logarithmically transforming MSE(Y) would make the variance like term MSE(Y) more stable.

Across all levels of the main effect the following conclusions, based on ln[MSE(Y)], were reached:

1. Shrunken weights were superior to MLR ($w^2 = 0.011$).
2. Validity weights were superior to unit weights ($w^2 = 0.020$).
3. The validity and unit weights were superior to the MLR and shrunken weights ($w^2 = 0.100$).
4. The structural techniques were superior to MLR, shrunken, unit, and validity weights ($w^2 = 0.072$).
5. Factoring both the predictors and criterion was superior to factoring only the predictor covariance matrix ($w^2 = 0.095$).
6. The communality estimate used (highest r verses R^2) does not appear to make much difference ($w^2 < 0.001$).
7. The principal component technique was superior to the factor analytic technique ($w^2 = 0.050$). However an interaction (crossover) occurred ($w^2 = 0.092$) between principal component verses factor analysis and factoring predictors verses factoring predictors and criterion. Factoring predictors only, with the principal component technique, was better than factoring both predictors and criterion. Factoring predictors only with the factor analytic technique, was worse than factoring with predictors and criterion.

Figure 1 shows the errors of prediction for the four levels of communality and five alternative weighting methods (MLR, shrunken, unit, validity weights, and principal components decomposing three factors of the predictors only [Pri-X-3]). With few exceptions the MLR technique was the poorest in prediction. The principal component technique was superior to the other four techniques. A linear interaction of communality with unit and validity weights was found ($w^2 = 0.028$). When the variables had low communality, unit weight was the best alternative. When the communality was high, unit weighting was the worst alternative. The

shrunken weights were consistently superior to MLR only. The validity weights were consistently superior to both MLR and shrunken weights. When the communality levels were high, medium, or wide, the validity weights were superior to the unit weights as well. The principal component weights were always the best for each level of communality, except for the low communality population, where unit weights were superior.

Figure 2 presents the above five alternative techniques for each level of N. As might be expected, as N increased the errors decreased, for all alternatives. The greatest relative gain was found for MLR. The principal component technique was superior to the other alternatives for each level of N. Shrunken weights were still consistently superior to MLR. However as N linearly increased the superiority of shrunken weights to MLR decreased ($w^2 = 0.012$). Only when N was very small were the unit weights viable. For every level of N, the validity weights outpredicted unit weights, this superiority increased linearly as N increased ($w^2 = 0.002$). When N was very large the amount of error in all methods presented (with the exclusion of unit weights) was approximately equivalent.

When enough factors were extracted, the F.A.-X method was the poorest, while the Pri-X method was the best of the four approaches. R^2 was used as the communality estimate here. In general, the effects due to highest r were identical to those of R^2 ($w^2 < 0.001$). If only one factor was extracted then the best method was F.A.-X, for each level of communality except the low communality population. A quadratic effect ($w^2 = 0.087$) was found for number of factors extracted across all population communality levels. With the exception of F.A.-X and the low communality population, a minimum amount of error was obtained, for each structural method when between two and four factors were extracted. With the exception of F.A.-X, as the communality decreased the other methods (Pri-X, Pri-X:Y, F.A.-X:Y) approached one another. The effects for the wide communality

level appear to be virtually identical to the effects of the medium communality population.

With the exception of F.A.-X (which in general was the poorest method, except when extracting one factor in a moderate or large sized sample), when N was small the minimum error was achieved by extracting between 2 and [illegible] factors. As N increased, the errors in prediction decreased for all methods and all levels of factors extracted. In addition, as N increased, extracting more factors had a less detrimental effect. That is, in large samples extracting more than three factors had little effect on errors, except for the Pri-X:Y method. As in the case of communality, if enough factors were extracted (greater than two), the principal component technique, decomposing only the predictors, dominated the other three alternative weighting systems for each level of N.

Figure 3 presents the logarithmically transformed errors of prediction for the four communality levels of the two and four factor populations, for each number of factors extracted of the structural technique. As communality increased, the errors decreased. As noted earlier, the medium and wide communality level populations had a similar pattern of effects across the levels of factors extracted. The two factor wide and medium populations had very similar means as well. The four factor wide communality apparently had more errors than the four factor medium communality populations. This may have been due to an artifact in the method used to generate the population covariance matrices. For the two factor populations, a minimum amount of errors in prediction was reached when two factors were extracted (i.e., extracting fewer or more factors generated more errors). For the four factor populations, the results were dependent on the communality level. For the four factor low communality level, a minimum amount of error was achieved when only one factor was extracted (extracting one factor in this population yielded less error than extracting zero factors by F.A.-X, [i.e.,

validity weights]). For the remaining three communality levels (high, medium, or wide), a relative minimum amount of error was achieved when extracting between three and four factors. The relative improvement, extracting the correct number of factors, was more pronounced as the communality level increased.

MSE(B) - Errors of Estimation of the Weights

One method used to access the dominance of the structural methods was to observe the relative frequency that a structural method (principal components extracting four factors of the predictor covariance matrix [Pri-X-4]) yielded less error than MLR. These results are presented in Table 4.

*** Insert Table 4 about here ***

These results were very similar to those presented in Table 2. When N was small (N = 16), using Pri-X-4, some improvements in estimation of the regression weights was found in every one of the 1600 independent samples. When N was moderate (N = 64), the principal component technique was superior to MLR in 96.44% of the observed samples. When N was large (N = 256), the principal component technique was still superior to MLR 83.75% of the time. For the medium or low communality populations, across all levels of sample size, the principal component technique was superior in virtually all of the 2400 samples. Only for the large N and wide communality population was the MLR as good as (if not better than) the principal component technique. That is, the principal component technique gave better prediction of the regression weights in all populations except for the large sample size in the wide communality population.

Table 4 presents the percentage of times principal component technique was superior to MLR to any degree. Table 5, below, presents the ratio of errors in MLR relative to the errors in the principal component technique (i.e., proportion of relative errors). As explained earlier, if this ratio were 1.0, the errors in MLR would be identical to the principal component technique. If this ratio were 2.0; for example, the MLR would have twice as many errors as the principal component technique.

*** Insert Table 5 about here ***

The results from this table indicate the clear superiority of the principal component technique over the MLR technique. When N was small a forty fold decrement in errors across all population levels of communality and factors was observed. With respect to the different levels of communality and population factors, when the sample size was small, improvements were found from a 27 fold improvement in prediction (two factor low communality) to a 65 fold improvement in prediction relative to MLR (four factor high communality). When N was moderate, principal components had between less than half (four factor wide communality) to less than one sixth (high or medium communality) the errors in MLR. For the moderate sample size, the average across all factor and communality levels was a 4.71 fold improvement over MLR. Even when N was very large, a 4.83 fold improvement was found (two and four factor medium communality and two factor wide communality).

With the exception of the wide communality population, with the large sample size, improvements from 91% to 63 times less error were found when using the principal component technique. When smaller sample sizes were used the amount of improvement increased.

The errors of estimation, MSE(B), were logarithmically transformed. Across all populations and N levels, the following was observed:

1. The shrunken weights were superior to MLR (w^2 = 0.014).
2. The validity weights were superior to unit weights (w^2 = 0.019).
3. The unit and validity weights were superior to the MLR and shrunken weights (w^2 = 0.214).
4. The structural method chosen (principal component or factor analytic weights decomposing predictors only or predictors and criterion) made a large difference (w^2 = 0.540). Pri-X was generally the best and F.A.-X was generally the worst alternative.
5. There was a linear (w^2 = 0.341), quadratic (w^2 = 0.068), and cubic (w^2 = 0.036) effect for number of factors extracted.
    1. A minimum amount of error was reached when between two to four factors were extracted.
    2. Extracting fewer or more factors increased errors.
    3. The optimum number of factors to extract was dependent upon which of the structural technique was used (w^2 = 0.262).

Figure 4 presents the ln[MSE(B)] for each of the four levels of communality by five alternative weighting procedures [MLR, shrunken, unit, validity, and principal components extracting three factors of the predictor covariance matrix). Similar to the results for ln[MSE(Y)], the shrunken weights were superior to MLR across each level of communality. The shrunken weights made the greatest improvements relative to MLR as the communality levels decreased (w^2 = 0.004). Unit weights were markedly superior to both MLR and shrunken weights across each level of communality. Validity weights were superior to unit weights in all populations except the low communality population, in which case a crossover occurred. In the low communality population unit weights were superior to validity weights. Finally, the Pri-X-3 method was superior to all other techniques, except in the wide communality population where the validity weights were best, followed closely by Pri-X-3.

Figure 5 presents the above five alternative methods for each level of N. Shrunken weights were superior to MLR across all levels of N (w^2 = 0.014), with the amount improvement related linearly to N (w^2 = 0.002). The greatest relative improvement was found for the smallest N; the least relative improvement was found in the largest N. Validity weights were, across all levels of N, superior to unit weights (w^2 = 0.019), although there was also an interaction linearly related to N (w^2 = 0.019). That is, validity weights were superior to unit weights when N was equal to or greater than 64, but poorer than unit weights when N was small. Finally, the principal component weights for each level of N was superior to all other methods.

Figures 6, 7 and 8 present the four structural procedures for each level N (16, 64, 256) by number of factors extracted. The errors decreased as N increased, across all methods. An interaction occured between sample size and the structural methods. When N was small, as presented in Figure 6, Pri-X-2 gave the least errors of estimation. F.A.-X was inferior to all other structural alternatives. For the other three sets of alternatives, extracting one factor yielded more error than extracting two factors. Extracting more than two factors also yielded more errors (i.e., there was a quadratic effect for factors extracted with minimum at two factors extracted) for these three alternative weighting systems. When N was 64, as presented in Figure 7, there was also a quadratic effect on the number of factors extracted for the other three alternative methods, the F.A.-X method was still inferior to the other alternatives. However, in this case the best method was factor analytic weights after decomposing both predictors and criterion (F.A.-X:Y), when extracting two factors. When more than two factors were extracted, Pri-X appeared to be the best alternative. When N was 256, as presented in Figure 8, F.A.-X:Y still yielded the greatest reduction on ln[MSE(B)], when extracting two factors. When more than three factors were

extracted, principal components decomposing the predictors only was superior. F.A.-X made improvements relative to the other structural alternatives in the large sample size. F.A.-X was better than the other alternatives only when one factor was extracted.

Figure 9 presents ln[MSE(B)] for each factor and N level and number of factors extracted across all the structural procedures. Again it may be noted that as N increased the amount of errors decreased. An interaction between the number of factors in the population and number of factors extracted (w^2 = 0.121) was found. For the two factor population, across all levels of N, a minimum number of errors at two factors extracted was observed. For the four factor populations, a minimum between three and four factors extracted was found. These minima, in the two or four factor populations, were dependent on the level of N. When N was 16, extracting one factor yielded less error. When N was large the minimum was found when the number of factors extracted was the same as in the population.

Considering the effect of the number of population factors on structural technique, across all four communality levels, F.A.-X was found to be among the poorest, except when one factor was extracted. The minimum errors were found when two factors were extracted. However the structural procedure, which was found to minimize errors varied with the communality level (i.e., for high communality - F.A.-X:Y was superior; for medium communality - Pri-X:Y was superior; for low communality - Pri-X was superior; and for wide communality - F.A.-X:Y was superior).

The difference between the factor analytic and principal component technique was linearly related to sample size (w^2 = 0.139). When N was small, the principal component techniques demonstrated relatively greater superiority to the factor analytic techniques. As noted earlier there was a tendency for the minimum to be

reached when two factors were extracted. A smaller (w^2 = 0.018) triple interaction also was found between the communality levels, a linear effect with sample size and the factor analytic - principal component difference. In this case, in the wide and high communality populations the factor analytic method in large samples tended to be superior to the principal component method. For the medium or low communality populations, or in the small or moderate sample sizes, the principal component method remained superior to the factor analytic method.

Summary and Conclusions

A number of weighting systems have been proposed in the literature for prediction of individual scores on a criterion and estimation of parameters. The two methods most frequently used are MLR and unit weights. In many ways MLR and unit weights are at opposite ends of a spectrum. MLR is a totally sample dependent statistic. With the exception of the very general assumptions in the model, i.e. the "classical normal regression model", MLR makes very weak assumptions about the data. In contrast, the unit weight technique makes quite strong assumptions about the data. Unit weights assume that 1) the investigator has a priori knowledge of the direction of each variable and 2) the optimal weights are equal to one another.

It might be expected than MLR would be poor if the data were weak (e.g., small N, unreliable data). If the data were very strong, (e.g., large N, reliable data) MLR would be expected to predict perfectly. Unit weights, no matter how much information was collected (except for correcting the sign of a weight or improvement in the standard deviations), would never improve. If the data were

weak (or non-existent) predictions could still be made by unit weights. Other weighting methods fall between unit and MLR weights, in terms of reliance on data and use of assumptions. The purpose of this study was to determine the most efficient of alternative weighting methods in estimating parameters and predicting individual scores. Forty different weighting systems (including MLR and unit weights) were examined. The results of this study showed, for most uses, these alternative weights invariably predicted better.

Under these conditions, vast improvements in prediction and estimation could be made over MLR and unit weights. These improvements were found in the overwhelming majority of observed samples (93.4%). The amount of improvement was as large as 65 times fewer errors. Unit weights were dominated by validity weights except for a few conditions (e.g., low communality levels and small N). The following conclusions were made:

1. MLR was invariably poorer than shrunken weights regardless of the underlying population or sample size. This was true for all criteria observed. On the other hand, use of shrunken weights will not affect (i.e., improve the power of) significance testing. These conclusions are similar to those attained by Gunst & Mason (1977) and Dempster, Schatzoff & Wermuth (1977).

2. Validity weights were superior to unit weights except when information was very poor (i.e., low communality or a large amount of error variance) and N was small. Claudy (1972) found similar results as well. However, in Claudy's study the validity weights became superior when N was quite large. In the present study only in the low communality populations were large samples needed to attain superiority of validity to unit weights.

3. When N was large and the factor model held, a structural method was better than all the other methods observed.

    1. For high, medium, or low communality, when N was large, for all criteria used, principal components decomposing the predictors [Pri-X] tended to be superior to factor analytic weights.

    2. When N was large and communality levels wide, factor analytic weights, decomposing the predictors and criterion matrix, was the best of all methods.

4. When N was small the structural methods were generally still superior to the other methods.

1. When N was small and high communality levels were present principal components were best for MSE(Y), while the validity weights were best for MSE(B). As earlier noted, these validity weights were the same as F.A.-X-O.

2. When the communality levels were low and N was small principal components decomposing predicitors only was best.

3. When the communality levels were wide (i.e., heterogeneous) and N was small factor analytic weights decomposing both predictor and criterion was best.

5. The number of factors a structural should extract depended on the number of factors in the population, sample size, and communality levels.

1. For a small number of factors in the population under or over factoring is more critical.

2. The number of factors was monotonically related to sample size.

3. When the communality level was high, a larger number of factors could be extracted.

As noted above, in point 3, the principal component technique did better than the factor analytic technique when a narrow band of communality was present (i.e., high, medium, or low). If the uniquenesses were all equal to one another (i.e., equal to k), then the factor analytic decomposition would be

14) $R = AA' + kI = V\Lambda V' + kI$

As the addition of a scalar matrix only affects the eigenroots, not the eigenvalues (Tatsuoka, 1971, 140), then Eqn. 28 is identical to

15) $R = V(\Lambda + kI)V'$

Equation 29 has a form identical to a principal component decomposition of R. Therefore, when the range of communalities levels was narrow, the principal component technique was a very useful approximation to the factor analytic model (with much fewer parameters to estimate), especially when little or poor information was available.

In sum, by the use of ancilliary information (e.g., factorial structure of the population, reliability of variables), an investigator may greatly maximize prediction and minimize errors. When information is very poor one composite (e.g., Pri-X-1, unit or validity weights) may be all the information that could successfully be extracted. When more information is available, extracting the full factorial model, which underlies the data, will extract all useful information.

An investigator may view the results of this study with some trepidation. After all, he might ask, "Who would consider using multiple regression with twelve predictors and only sixteen subjects?" Doing multiple regression with so few subjects violates all current rules of practice. This study has reaffirmed that multiple regression is not appropriate with so few subjects. However, Pri-X-3, may successfully be used with sixteen subjects. In addition, MLR has often been successfully used with 61 subjects. By interpolation, the use of MLR with 61 subjects is equivalent to the use of Pri-X-3 with sixteen subjects. In many settings, obtaining subjects is not only difficult but often expensive. In addition, the existence of pre-existing groups often forces the investigator to work within separate small groups only. Therefore, one important use of the results of this study becomes apparent. It is believed that the results of this study show strong evidence in favor of using structural methods and other alternative weighting methods even with very small sample sizes, perhaps even samples of sixteen subjects. However, even when large samples were available, gains were frequently realized as well.

The findings presented in this study were totally dependent on the characteristics of the population. If the true population factor loadings were changed (e.g., allowing for either greater or lesser correlations amoung factors, fewer or more factors, greater or lesser adherance to a positive manifold); if the

number of predictors were changed; if large departures from the factor model were allowed; or if non-normal distributions were utilized, then the conclusions might be radically altered. Future research should be undertaken to see the effects of the above manipulations. In addition, if would be quite useful to obtain the confidence intervals for the alternative weighting methods, especially for the structural methods. Lastly, other alternative weighting methods (e.g., preliminary testing, ridge regression) might be considered.

References

Burket, G. R., A study of reduced rank models for multiple predictions. Psychometric Monographs, 1964, 12.

Claudy, J. G., A comparison of five variable weighting procedures. Educational and Psychological Measurement, 1972, 32, 311-322.

Dawes, R. M., & Corrigan, B., Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.

Dempster, A. P., Schatzoff, M., & Wermuth, N., A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association, 1977, 72, 77-91.

Dorans, N. & Drasgow, F., Alternative weighting schemes for linear prediction. Organizational Behavior and Human Performance, 1978, 21, 316-345

Einhorn, H. J. & Hogarth, R. M., Unit weighting schemes for decision making. Organizational Behavior and Human Performance, 1975, 13, 171-192.

Fleishman, A. I., Confidence intervals for correlation ratios. In press, Educational and Psychological Measurements.

Fleishman, A. I., Investigation and comparison of Alternative Weighting Systems, doctoral dissertation, University of Illinois, 1978.

Gunst, R. F., & Mason, R. L., Biased estimation in regression: an evaluation using mean squared error. Journal of the American Statistical Association, 1977, 72, 616-628.

Herzberg. P. A., The parameters of cross-validation. Psychometrika Monograph Supplement, 1969, 16.

Horst, P., The Prediction of Personal Adjustment. New York: Social Science Research Council, 1941.

Joreskog, K. G., On the statistical treatment of residuals in factor analysis. Psychometrika, 1962, 27, 335-354.

Kmenta, J., Elements of Econometrics. New York: Macmillian Co., 1971.

Laughlin, J. E., Comment on "estimating coefficients in linear models: it don't make no nevermind". Psychological Bulletin, 1978, 85, 247-253.

Lawley, D. N., & Maxwell, A. E., Regression and factor analysis. Biometrika, 1973, 60, 331-338.

Lawshe, C. H. & Schucker, R. E., The relative efficiency of four test weighting methods in multiple prediction. Educational and Psychological Measurement, 1959, 19, 103-114.

Leiman, J. M., The calculation of regression weights from common factor loadings, Doctoral dissertation, University of Washington, 1951.

Pruzek, R. M., Sherry, J. R., & Huba, M. E., Improving least squares regression estimators through matrix factoring. Paper presented at Psychometric Society, April, 1975, Iowa City, Iowa.

Pruzek, R. M. & Frederick, B. C. Weighting predictors in linear models: alternatives to least squares and limitations of equal weights. Psychological Bulletin, 1978, 85, 254-267.

Sampson, A. R., A tale of two regressions. Journal of the American Statistical Association, 1974, 69, 682-689.

Schmidt, F. L., The relative efficiency of regression and simple unit predictor weights in applied differential psychology. Educational and Psychological Measurement, 1971, 31, 699-714.

Schmidt, F. L., The reliability of differences between linear regression weights in applied differential psychology. Educational and Psychological Measurement, 1972, 32, 879-886.

Sclove, S. L., Improved estimators for coefficients in linear regression. Journal of the American Statistical Association, 1968, 63, 596-606.

Scott, Jr., J. T., Factor analysis and regression. Econometrica, 1966, 34, 552-562.

Scott, Jr., J. T., Notes and comments: factor analysis regression revisited. Econometrica, 1969, 37, 719.

Scott, Jr., J. T., Combining regression and factor analysis for use in agricultural economics research. Paper presented at Southern Agricultural Economics Association, February 1976, Mobile, Alabama.

Scott, Jr., J. T., & Fleishman, A. I., Statistical analysis of the goodness of classical factor analysis regression (CFAR). Bulletin, Department of Agricultural Economics, University of Illinois, Urbana, Illinois, February, 1977.

Scott, Jr., J. T., & Fleishman, A. I., Classical factor analysis regression (CFAR) and its statistical properties. Paper presented at Western Economics Association, June, 1976, San Francisco, California.

Stein, C., Multiple regression, in I. Olkin (Ed.) Contributions to Probability and Statistics. California: Stanford University Press, 1960, 424-443.

Tucker, L. R, Koopman, R. F., & Linn, R. L., Evaluation of Factor analytic research procedures by means of simulated correlation matrices. Psychometrika, 1969, 34, 421-459.

Wainer, H., Estimating coefficients in linear models: it don't make no nevermind. Psychological Bulletin, 1976, 83, 213-217.

Wesman, A. G. & Bennett, G. K., Multiple regression vs. simple addition of scores in prediction of college grades. Educational and Psychological Measurement, 1959, 19, 243-246.

Wijsman, R. A., Application of a certain representation of the Wishart matrix. Annals of Mathematical Statistics, 1959, 30, 597-601.

Winer, B. J., Statistics and data analysis: Trading bias for reduced mean square error. Annual Review of Psychology, 1978, 29, 647-681.

Wishart, J , The generalized product moment distribution in samples from a normal multivariate population. Biometrika, 1928, 20, 32-52.

Table 1

Regression Weights and R^2 for Population 1

	Two Factors				Four Factors			
	Communalities							
	High	Medium	Low	Wide	High	Medium	Low	Wide
B_1	-0.0283	-0.0111	-0.0005	-0.0277	-0.0570	-0.0442	-0.0190	-0.0158
B_2	0.1071	0.1068	0.0593	0.0284	0.1346	0.1230	0.0608	0.0381
B_3	0.1129	0.0850	0.0526	0.0549	-0.0732	-0.0275	-0.0018	-0.0681
B_4	0.2457	0.1575	0.0904	0.1959	0.3206	0.1945	0.0999	0.2408
B_5	-0.1121	-0.0930	-0.0467	-0.0450	0.1251	0.0966	0.0408	0.0403
B_6	-0.0035	0.0077	0.0097	-0.0141	-0.0533	-0.0265	-0.0046	-0.0336
B_7	0.0528	0.0627	0.0312	0.0058	0.0337	0.0422	0.0216	0.0033
B_8	0.1609	0.1416	0.0764	0.0578	0.2080	0.1766	0.0846	0.0641
B_9	-0.0537	-0.0443	-0.0193	-0.0197	0.0184	0.0062	-0.0003	0.0112
B_{10}	0.0881	0.0896	0.0510	0.0222	0.0554	0.0582	0.0338	0.0134
B_{11}	0.1424	0.1262	0.0718	0.0503	0.1953	0.1516	0.0732	0.0775
B_{12}	-0.0755	-0.0688	-0.0321	-0.0225	-0.1626	-0.1444	-0.0598	-0.0397
R^2	0.6806	0.3474	0.0647	0.1195	0.6626	0.3156	0.0513	0.1123
R	0.8250	0.5894	0.2544	0.3457	0.8140	0.5618	0.2265	0.3350
Det.	$.3687x10^{-6}$	$.6113x10^{-2}$	.3626	$.2494x10^{-2}$	$.1322x10^{-4}$	$.2633x10^{-1}$	.5247	$.1679x10^{-1}$

Table 2

Percentage of Times Pri-X-4 Yielded Less Error than MLR in MSE(Y)

		N		
Factors	Communality	16	64	256
2	High	100	100	97
2	Medium	100	100	100
2	Low	100	100	100
2	Wide	100	98.5	65
4	High	100	100	84
4	Medium	100	99.5	98.5
4	Low	100	99.5	99.5
4	Wide	100	82.5	49.5
Mean (across factors and communality)		100	97.5	86.69

Table 3

Ratio MSE(MLR)/MSE(Pri-X-4) in MSE(Y)

		N		
Factors	Communality	16	64	256
2	High	6.21	1.16	1.03
2	Medium	5.54	1.17	1.03
2	Low	5.59	1.18	1.03
2	Wide	5.18	1.17	1.02
4	High	5.11	1.15	1.02
4	Medium	4.39	1.15	1.03
4	Low	5.19	1.17	1.03
4	Wide	5.43	1.11	1.00
Mean (across factors and communality)		5.28	1.16	1.02

Table 4

Percentage of Times Pri-X-4 Yielded Less Errors than MLR in MSE(B)

		N		
Factors	Communality	16	64	256
2	High	100	100	97
2	Medium	100	100	100
2	Low	100	100	100
2	Wide	100	94	51.5
4	High	100	100	79.5
4	Medium	100	100	99.5
4	Low	100	100	100
4	Wide	100	77.5	42.5
Mean (across Factors and Communality)		100	96.44	83.75

Table 5

Ratio MSE(MLR)/MSE(Pri-X-4) in MSE(B)

		N		
Factors	Communality	16	64	256
2	High	63.85	6.33	3.06
2	Medium	46.92	6.47	4.83
2	Low	35.70	5.77	4.11
2	Wide	44.08	4.59	1.36
4	High	64.67	6.65	1.91
4	Medium	30.76	6.34	4.39
4	Low	27.47	4.24	3.33
4	Wide	31.69	2.39	0.65
Mean (across factors and communality)		40.11	4.71	1.85

Figure Captions

Figure 1: Errors of Prediction for each Communality Level for Selected Methods

Figure 2: Errors of Prediction for each Level N for Selected Methods

Figure 3: Errors of Prediction for each Communality Level and Number of Population Factor for each of the Structural Methods

Figure 4: Errors of Estimation for each Communality Level for Selected Methods

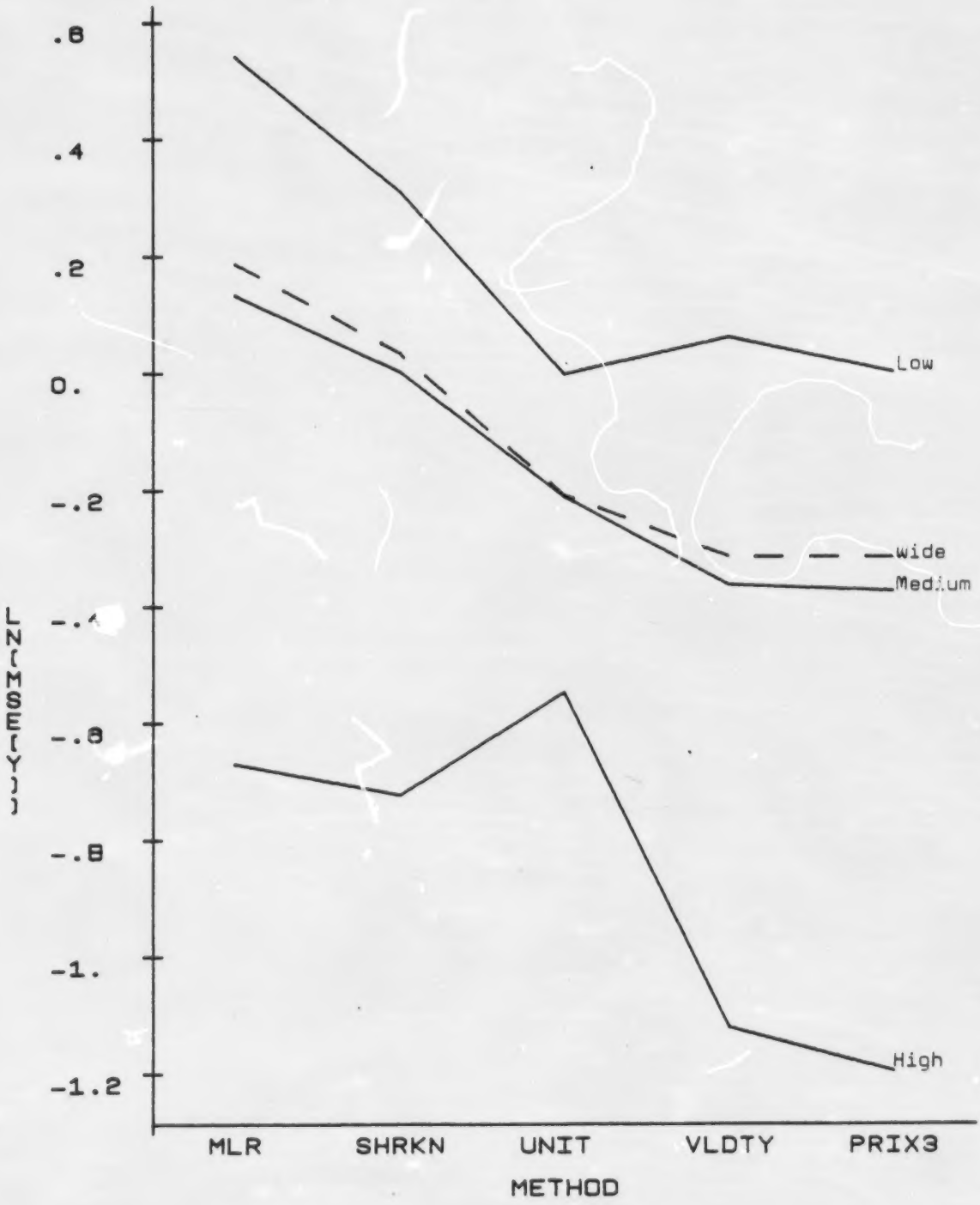
Figure 5: Errors of Estimation for each Level N for Selected Methods

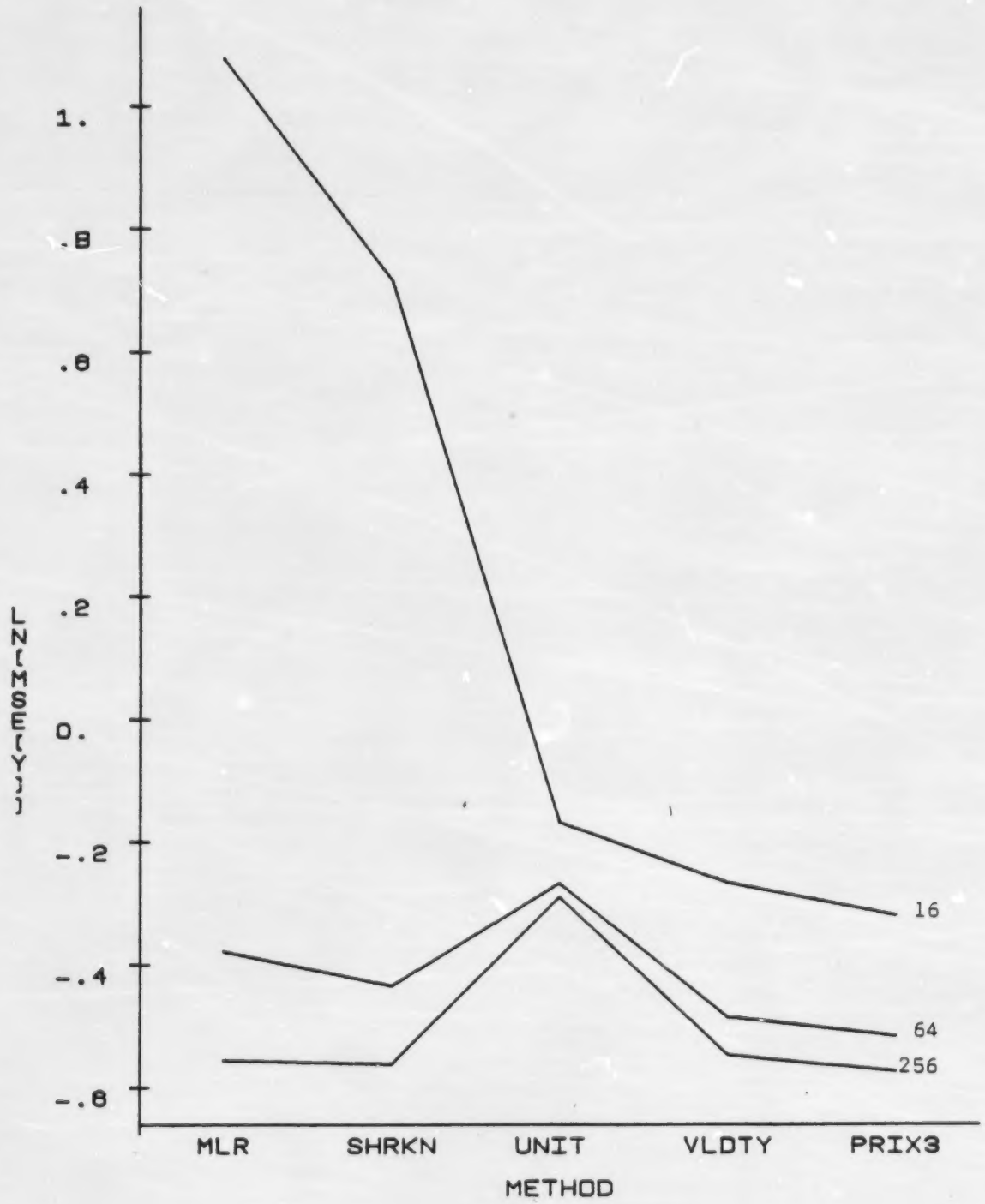
Figure 6: Errors of Estimation for N = 16 for the Structural Methods

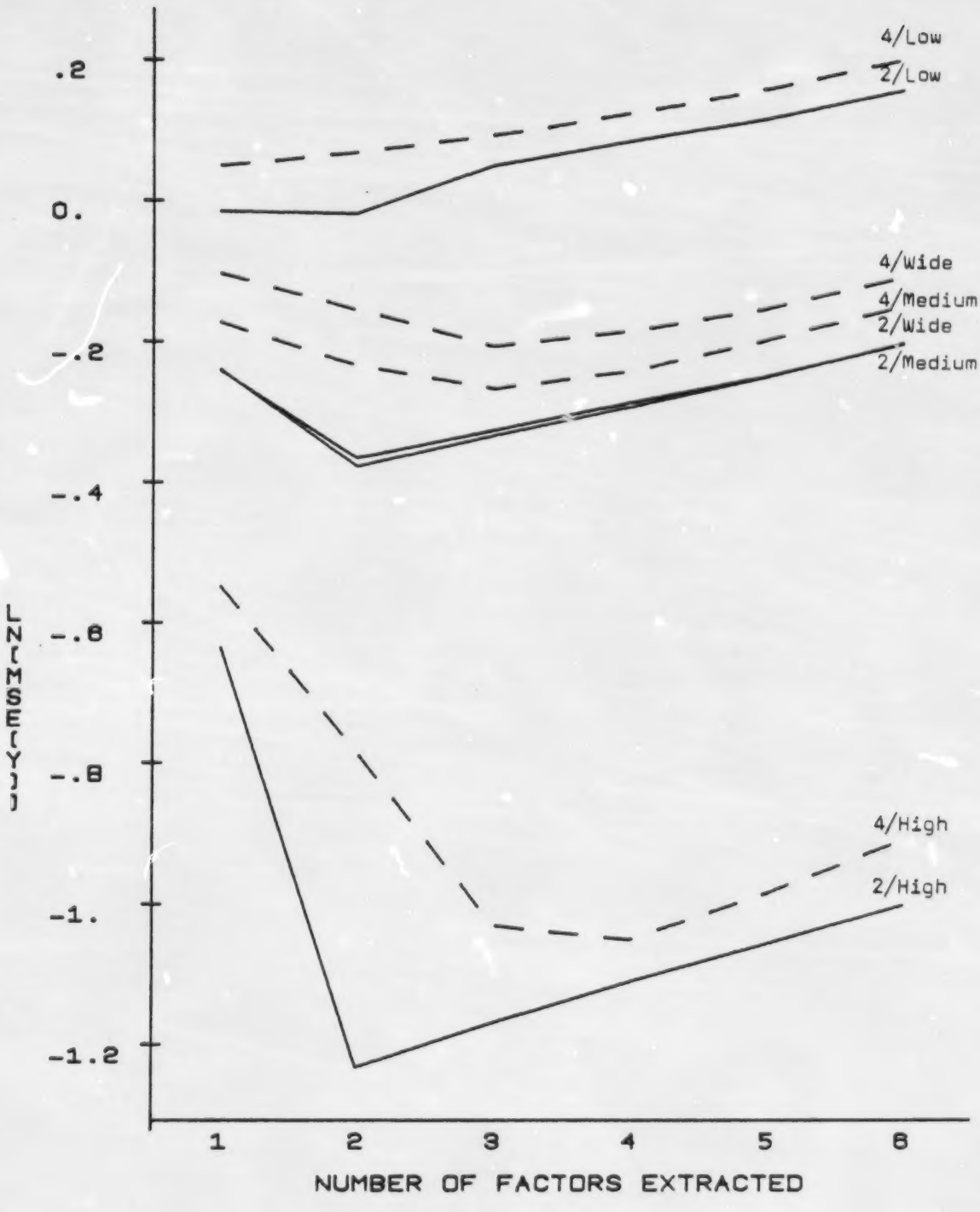
Figure 7: Errors of Estimation for N = 64 for the Structural Methods

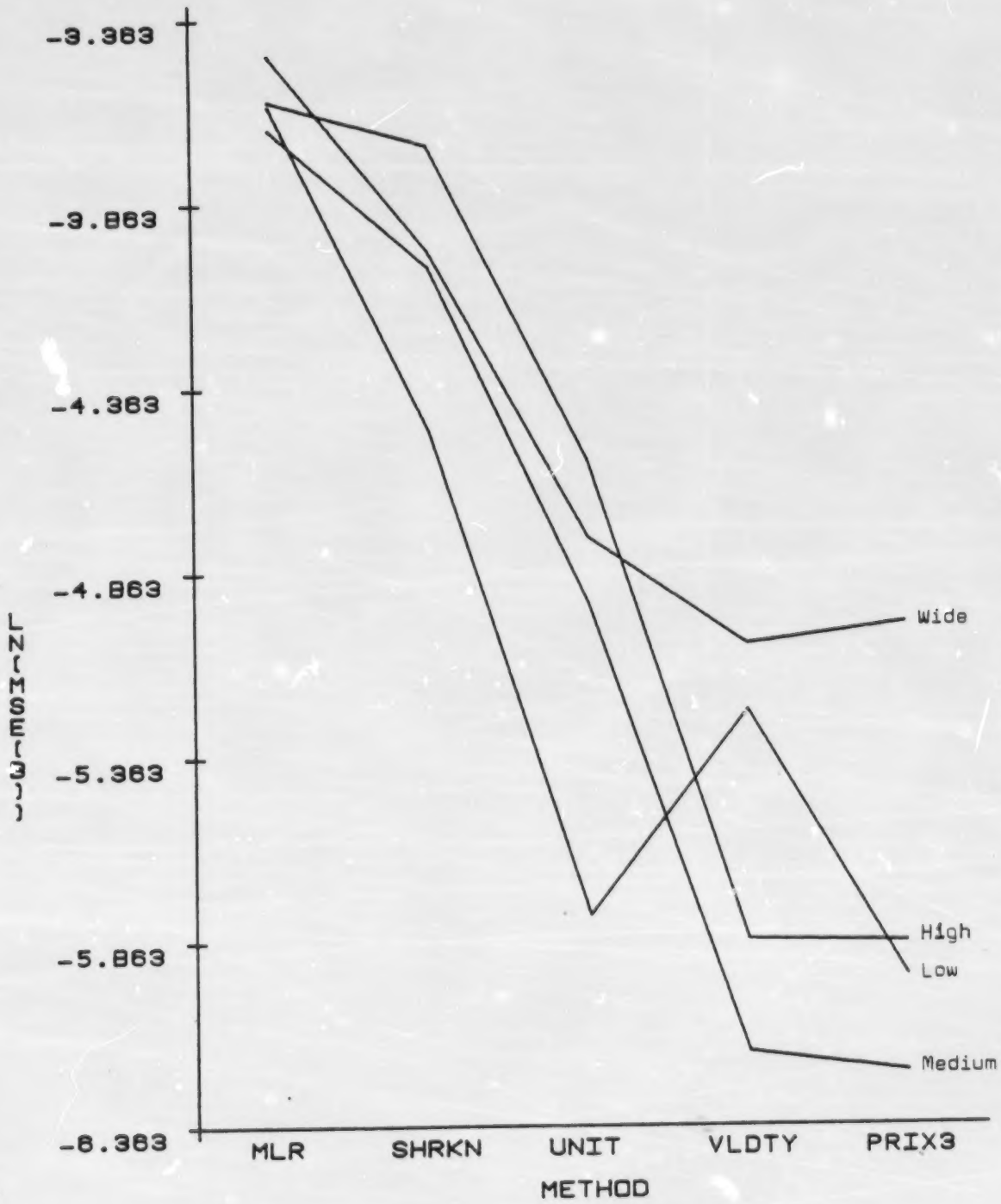
Figure 8: Errors of Estimation for N = 256 for the Structural Methods

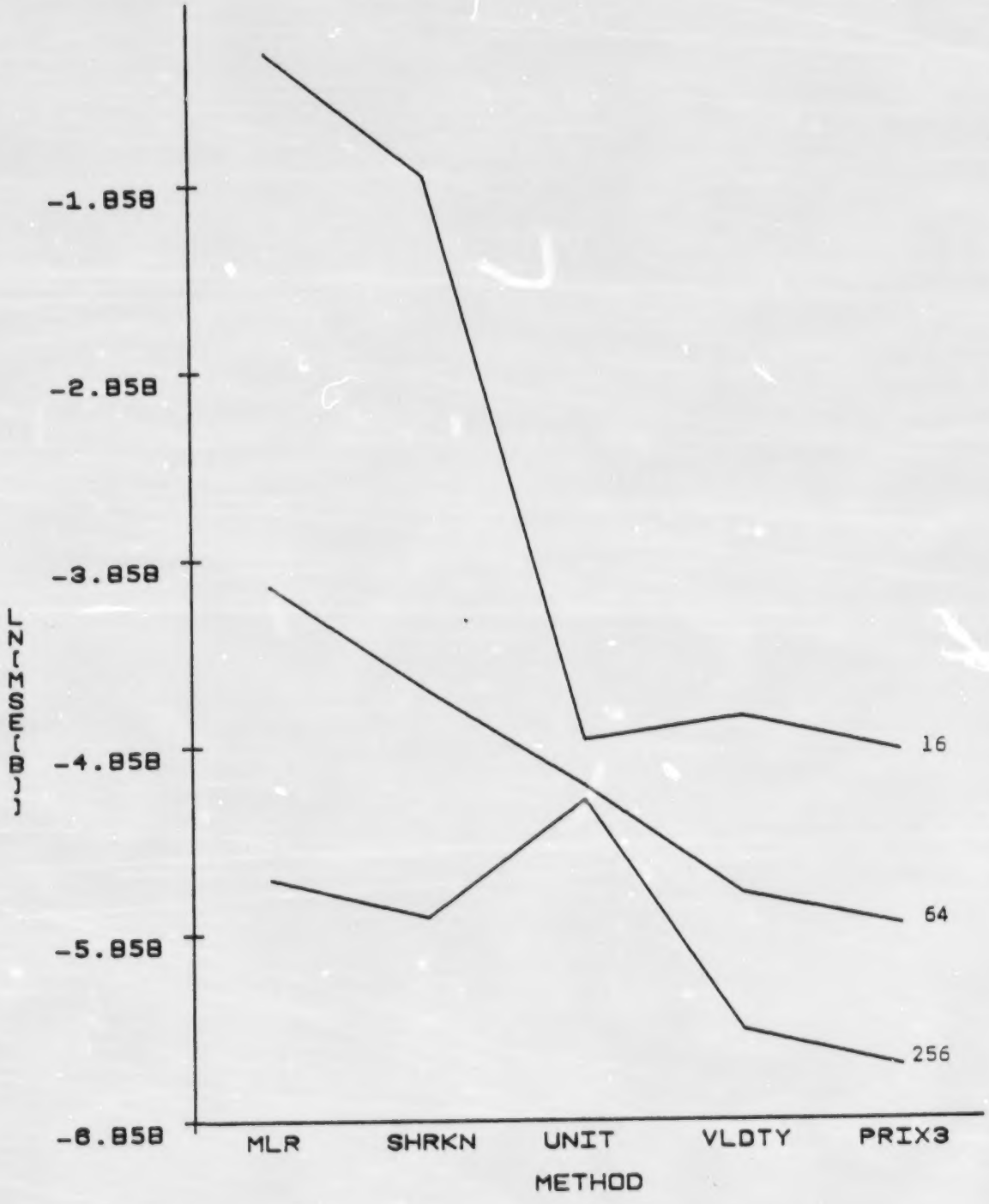
Figure 9: Errors of Estimation for Number of Population Factors and Level of N for the Structural Methods

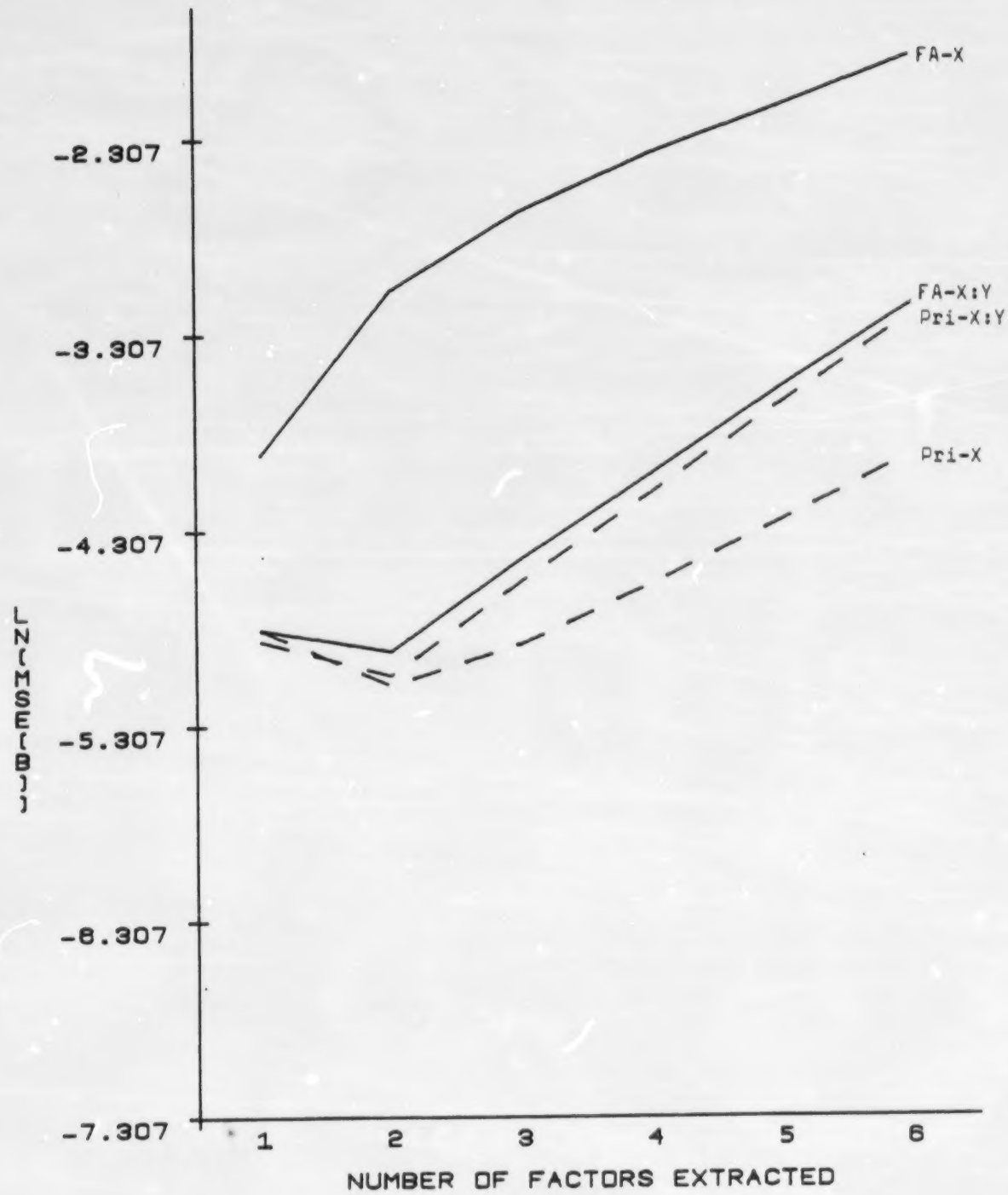

.6
.4
.2
0.
-.2
-.4
-.6
-.8
-1.
-1.2
LN(MSE(Y))
Low
wide
Medium
High
MLR
SHRKN
UNIT
VLDTY
PRIX3
METHOD

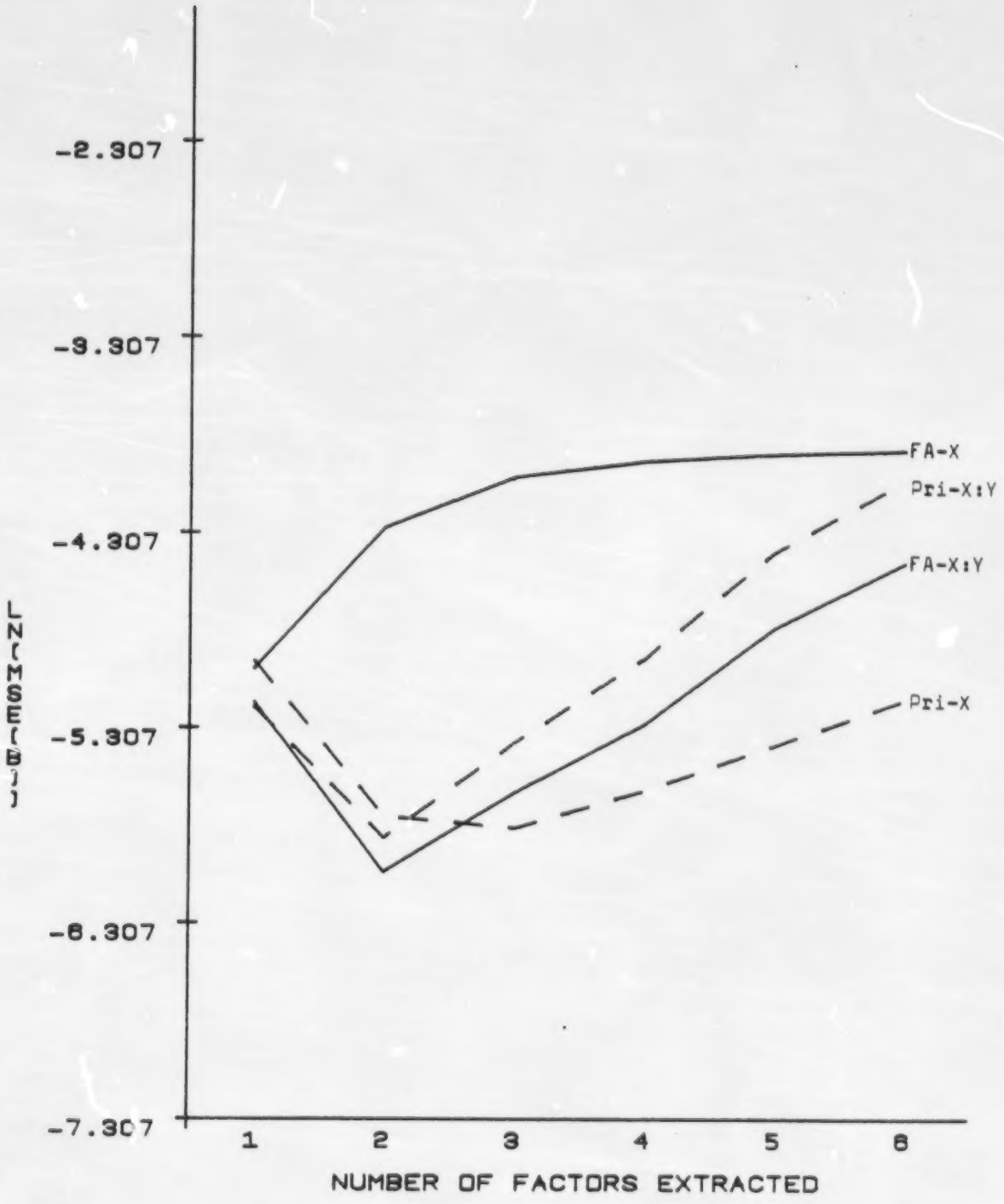

1.
.8
.6
.4
.2
0.
-.2
-.4
-.6
LN(MSE(Y))
16
64
256
MLR
SHRKN
UNIT
VLDTY
PRIX3
METHOD

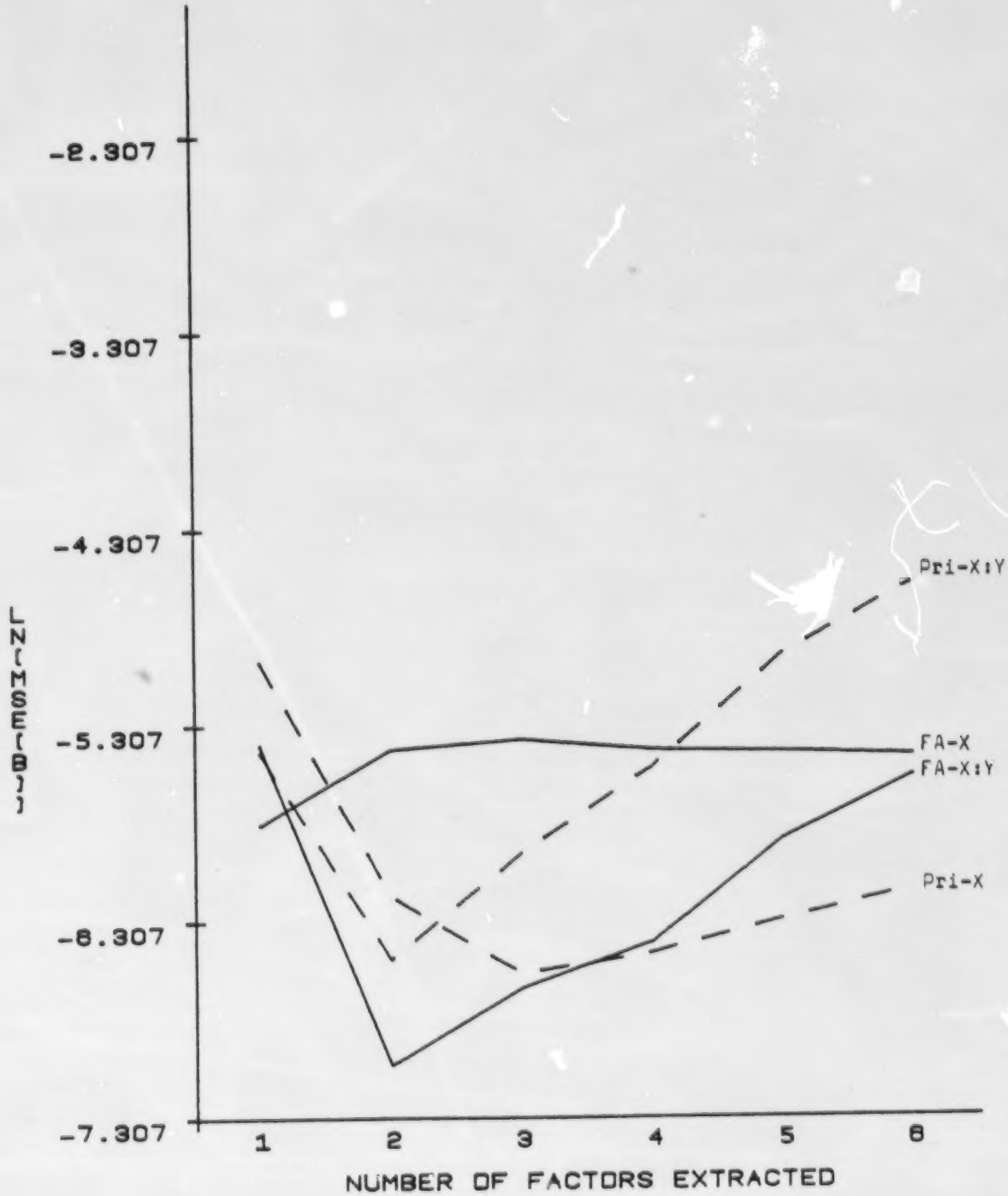

4/Low
2/Low
4/Wide
4/Medium
2/Wide
2/Medium
4/High
2/High
.2
0.
-.2
-.4
-.6
-.8
-1.
-1.2
LN[MSE(Yj)]
1
2
3
4
5
6
NUMBER OF FACTORS EXTRACTED

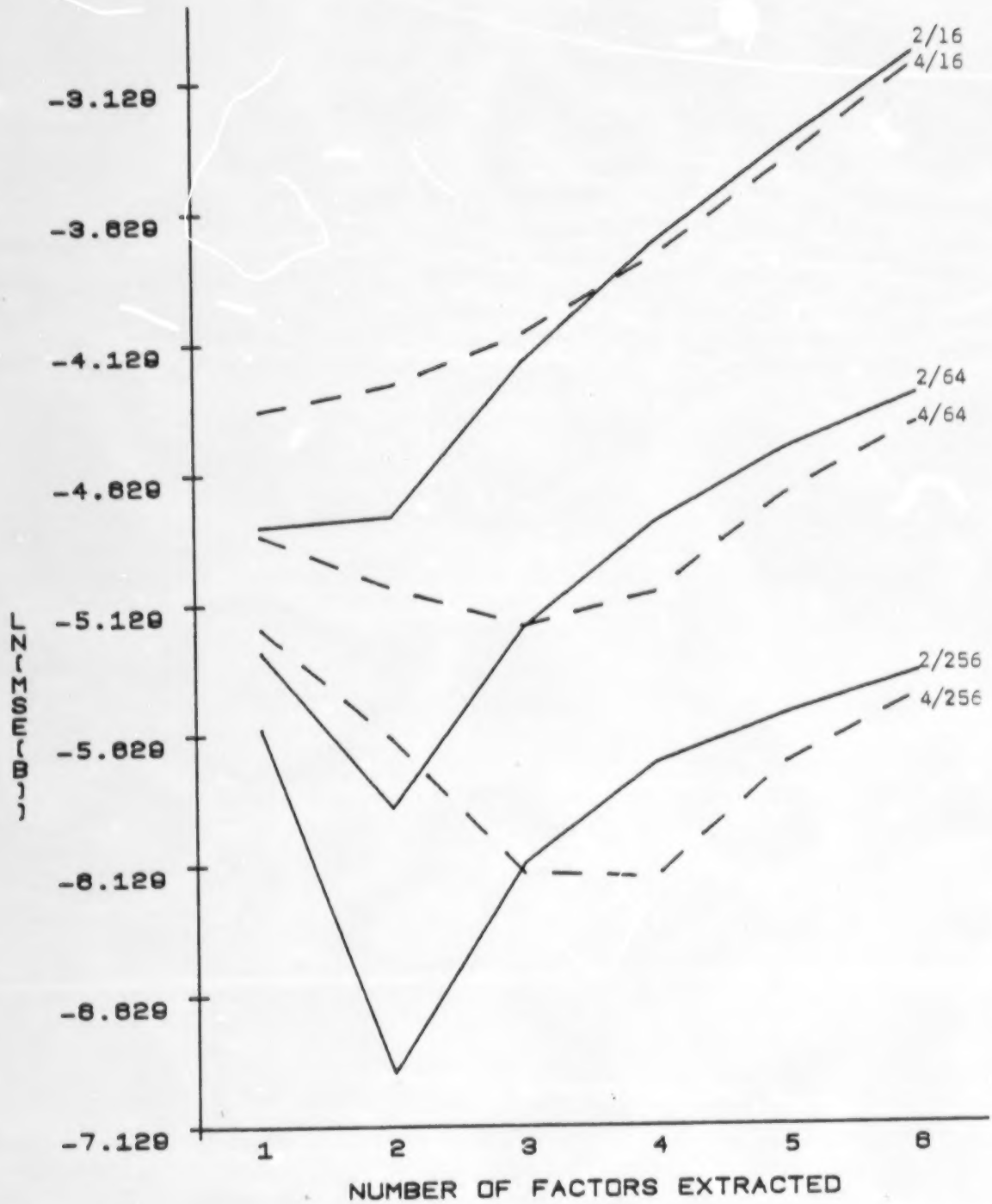

-3.383
-3.863
-4.383
-4.863
-5.383
-5.863
-6.383
LN(MSE(3))
MLR
SHRKN
UNIT
VLDTY
PRIX3
METHOD
Wide
High
Low
Medium


-1.858
-2.858
-3.858
-4.858
-5.858
-6.858
LN(MSE(B))
16
64
256
MLR
SHRKN
UNIT
VLDTY
PRIX3
METHOD


FA-X
FA-X:Y
Pri-X:Y
Pri-X
-2.307
-3.307
-4.307
-5.307
-6.307
-7.307
LN(MSE(Bj))
1
2
3
4
5
6
NUMBER OF FACTORS EXTRACTED


-2.307
-3.307
-4.307
-5.307
-6.307
-7.307
LN(MSE(B))
1
2
3
4
5
6
NUMBER OF FACTORS EXTRACTED
FA-X
Pri-X:Y
FA-X:Y
Pri-X


-2.307
-3.307
-4.307
-5.307
-6.307
-7.307
LN(MSE(B))
1
2
3
4
5
6
NUMBER OF FACTORS EXTRACTED
Pri-X:Y
FA-X
FA-X:Y
Pri-X


2/16
4/16
2/64
4/64
2/256
4/256
-3.128
-3.628
-4.128
-4.628
-5.128
-5.628
-6.128
-6.628
-7.128
LN(MSE(B))
1
2
3
4
5
6
NUMBER OF FACTORS EXTRACTED